WHAT IS CLAIMED IS:

1. A cDNA microarray data correction system for correcting global and local distortions of microarray data more precisely and correcting measurement errors caused by a difference in sensitivity between fluorescent dyes, comprising:

an input device for inputting previously-adjusted gene expression intensity data, considering flag information indicating a removal of background noise and reliability of each spot;

a data standardization means for standardizing the gene expression intensity data by using grid-by-grid order statistics for the input gene expression intensity data and for transmitting the standardized gene expression intensity data;

first correction means for estimating a distortion depending on a spot position on grid coordinates for the standardized gene expression intensity data by a nonparametric smoothing method and for transmitting first corrected gene expression intensity data whose distortion has been corrected; and

second correction means for performing an S-D transformation for the first corrected gene expression intensity data, for estimating a potential distortion caused by a difference in sensitivity between the fluorescent dyes in the gene expression intensity data by the nonparametric smoothing method, and for transmitting second corrected gene expression intensity data whose

distortion caus d by the difference in sensitivity between the fluorescent dyes has b en corrected; and

an output device for outputting the second corrected gene expression intensity data.

2. The cDNA microarray data correction system according to claim 1, further comprising S-D transformation means for quantifying the distortion of the gene expression intensity data in an arbitrary stage and for visualizing it on an S-D plot.

3. The cDNA microarray data correction system according to claim 1 or 2, wherein the order statistics are represented by the following EQ12 (where $w_{ij}^{k}(c)$ is the standardized gene expression intensity data, $y_{ij}^{k}(c)$ is gene expression intensity data of all spots obtained in a channel, and $L_k(c)$ and $M_k(c)$ indicate 25 and 50 percent points of the gene expression intensity data obtained in channel c in grid k, respectively):

$$w_{ij}^{k}(c)=\frac{y_{ij}^{k}(c) - L_k(c)}{M_k(c) - L_k(c)},\quad c=1,2,\ i=1,\cdots,I,\ j=1,\cdots,J,\ k=1,\cdots,K. \tag{12}$$

4. The cDNA microarray data correction system according to claim 1 or 2, wherein the order statistics are represented by the following EQ13 (where $w_{ij}^{k}(c)$ is the standardized gene expression intensity data, $y_{ij}^{k}(c)$ is gene expression intensity data of all spots obtained in a channel, and $A_k(c)$, $L_k(c)$ and $M_k(c)$ indicate 35, 10 and 90 percent points of the gene expression intensity data

obtained in channel c in grid k, respectively):

$$w_{ij}^{k}(c) = \frac{y_{ij}^{k}(c) - A_k(c)}{M_k(c) - L_k(c)},$$

$$c = 1,2, i = 1,\cdots,I, j = 1,\cdots,J, k = 1,\cdots,K. \quad (13)$$

5. The cDNA microarray data correction system according to claim 3 or 4, wherein said data standardization means determines whether the gene expression intensity data of all spots obtained in at least two gene expression intensity data channels has been standardized and continues it until the gene expression intensity data of all spots has been standardized.

6. The cDNA microarray data correction system according to claim 1, wherein the standardized gene expression intensity data is represented by a sum of a true gene intensity and a distortion depending on the spot position.

7. The cDNA microarray data correction system according to claim 1, wherein said first correction means describes the distortion depending on the spot position by means of a nonparametric regression model represented by a regression relation of distortions with an x-axis, a y-axis, and an interaction of the x- and y-axes ($\alpha_k^{(c)}(i)$, $\beta_k^{(c)}(j)$, and $\gamma_k^{(c)}((i - m_i)(j - m_j))$, respectively) and estimates the distortion depending on the spot position ($\xi_{ij}^{k}(c)$) by the nonparametric smoothing method represented by the following EQ14:

$$\hat{\xi}^k_{ij}(c)=\hat{\alpha}^{(c)}_k(i)+\hat{\beta}^{(c)}_k(j)+\hat{\gamma}^{(c)}_k((i-m_i)(j-m_j)),\ c=1,2,\ i=1,\cdots,I,\ j=1,\cdots,J. \quad (14)$$

8. The cDNA microarray data correction system according to claim 7, wherein the distortion depending on the spot position is corrected according to the following EQ15 (where $\hat{z}^k_{ij}(c)$ is corrected true gene expression intensity data):

$$\hat{z}^k_{ij}(c) = w^k_{ij}(c) - \hat{\xi}^k_{ij}(c) \quad (15)$$

9. The cDNA microarray data correction system according to claim 8, wherein the S-D transformation in said second correction means is performed according to the following EQ16:

$$u^k_{ij}= \hat{z}^k_{ij}(1) + \hat{z}^k_{ij}(2)$$
$$v^k_{ij}= \hat{z}^k_{ij}(1) - \hat{z}^k_{ij}(2) \quad (16)$$

10. The cDNA microarray data correction system according to claim 9, wherein said second correction means describes the distortion by means of a nonparametric regression model represented by the following EQ17, estimates a measurement error caused by the difference in sensitivity between the fluorescent dyes by a nonparametric smoothing method represented by the following EQ18 and EQ19, and corrects the error:

$$v^k_{ij} = \Phi(u^k_{ij}) + \varepsilon^k_{ij},\ \varepsilon^k_{ij} \sim N(0, v^2) \quad (17)$$

$$\eta_{ij}^{k} = v_{ij}^{k} - \hat{\phi}\left(u_{ij}^{k}\right) \tag{18}$$

$$\hat{y}_{ij}^{k}(1) = \frac{1}{2}\left(u_{ij}^{k} + \eta_{ij}^{k}\right)$$
$$\hat{y}_{ij}^{k}(2) = \frac{1}{2}\left(u_{ij}^{k} - \eta_{ij}^{k}\right) \tag{19}$$

11. The cDNA microarray data correction system according to claim 1, wherein, supposing that a probability of gene expression is lower than 0.5, it is assumed for the correction that the fluorescence intensity detected at more than half of the spots within each grid indicates a background noise or a systematic error.

12. The cDNA microarray data correction system according to claim 11, wherein, supposing that $L_k(c)$ and $M_k(c)$ indicate 25 and 50 percent points of the fluorescence intensity obtained in at least two gene expression intensity data channels in a grid, it is further assumed for the correction that $L_k(c)$ and $M_k(c) - L_k(c)$ are equal among the grids and teh channels on condition that most genes are in a non-expression state and that a distribution of 50 percent point or lower of the fluorescence intensity is common to all grids and channels.

13. A cDNA microarray data correction method of correcting global and local distortions of microarray data more precisely and correcting measurement errors caused by a difference in sensitivity between fluorescent dyes,

comprising the steps of:

inputting previously-adjusted gene expression intensity data, considering flag information indicating a removal of background noise and reliability of each spot;

standardizing the gene expression intensity data by using grid-by-grid order statistics for the input gene expression intensity data on condition that most genes are in a non-expression state;

outputting the standardized gene expression intensity data;

estimating a distortion depending on the spot position on grid coordinates for the standardized gene expression intensity data by a nonparametric smoothing method and correcting the data distortion depending on the spot position;

outputting the first corrected gene expression intensity data whose distortion depending on the spot position has been corrected;

performing an S-D transformation for the first corrected gene expression intensity data, estimating a potential distortion caused by a difference in sensitivity between the fluorescent dyes in the gene expression intensity data by the nonparametric smoothing method, and correcting the distortion caused by the difference in sensitivity between the fluorescent dyes; and

outputting the second corrected gene expression intensity data whose distortion caused by the difference in sensitivity between the fluorescent dyes has been corrected.

14. The cDNA microarray data correction method according to claim 13, further comprising a step of quantifying the distortion of the gene expression intensity data in an arbitrary stage and visualizing it on an S-D plot.

15. The cDNA microarray data correction method according to claim 13 or 14, wherein the order statistics are represented by the following EQ20 (where $w_{ij}^{k}(c)$ is the standardized gene expression intensity data, $y_{ij}^{k}(c)$ is gene expression intensity data of all spots obtained in a channel, and $L_k(c)$ and $M_k(c)$ indicate 25 and 50 percent points of the gene expression intensity data obtained in channel c in grid k, respectively):

$$w_{ij}^{k}(c)=\frac{y_{ij}^{k}(c)-L_k(c)}{M_k(c)-L_k(c)},\quad c=1,2,\ i=1,\cdots,I,\ j=1,\cdots,J,\ k=1,\cdots,K. \tag{20}$$

16. The cDNA microarray data correction method according to claim 13 or 14, wherein the order statistics are represented by the following EQ21 (where $w_{ij}^{k}(c)$ is the standardized gene expression intensity data, $y_{ij}^{k}(c)$ is gene expression intensity data of all spots obtained in a channel, and $A_k(c)$, $L_k(c)$ and $M_k(c)$ indicate 35, 10 and 90 percent points of the gene expression intensity data obtained in channel c in grid k, respectively):

$$w_{ij}^{k}(c)=\frac{y_{ij}^{k}(c)-A_k(c)}{M_k(c)-L_k(c)},$$

$$c=1,2,i=1,\cdots,I,j=1,\cdots,J,k=1,\cdots,K. \tag{21}$$

17. The cDNA microarray data correction method according to claim 15 or 16, wherein, in the step of standardizing the data, it is determined whether the gene expression intensity data of all spots obtained in at least two gene expression intensity data channels have been standardized and it is continued until the gene expression intensity data of all spots have been standardized.

18. The cDNA microarray data correction method according to claim 17, wherein the standardized gene expression intensity data is represented by a sum of a true gene intensity and a distortion depending on the spot position.

19. The cDNA microarray data correction method according to claim 13, wherein, in the step of correcting the data distortion depending on the spot position, the distortion depending on the spot position is described by means of a nonparametric regression model represented by a regression relation of distortions with an x-axis, a y-axis, and an interaction of the x- and y-axes ($\alpha_k^{(c)}(i)$, $\beta_k^{(c)}(j)$, and $\gamma_k^{(c)}((i - m_i)(j - m_j))$, respectively) and the distortion depending on the spot position ($\xi_{ij}^k(c)$) is estimated by the nonparametric smoothing method represented by the following EQ22:

$$\hat{\xi}_{ij}^k(c) = \hat{\alpha}_k^{(c)}(i) + \hat{\beta}_k^{(c)}(j) + \hat{\gamma}_k^{(c)}((i - m_i)(j - m_j)),$$
$$c = 1,2,\ i = 1, \cdots, I,\ j = 1, \cdots, J. \quad (22)$$

20. The cDNA microarray data correction method according to claim 19, wherein the distortion depending on

the spot position is corrected according to the following EQ23 (where $\hat{z}_{ij}^{k}(c)$ is corrected true gene expression intensity data):

$$\hat{z}_{ij}^{k}(c) = w_{ij}^{k}(c) - \hat{\xi}_{ij}^{k}(c) \quad (23)$$

21. The cDNA microarray data correction method according to claim 19, wherein the S-D transformation in the step of correcting the distortion caused by the difference in sensitivity between fluorescent dyes is performed according to the following EQ24:

$$u_{ij}^{k} = \hat{z}_{ij}^{k}(1) + \hat{z}_{ij}^{k}(2)$$
$$v_{ij}^{k} = \hat{z}_{ij}^{k}(1) - \hat{z}_{ij}^{k}(2) \quad (24)$$

22. The cDNA microarray data correction method according to claim 20, wherein, in the step of correcting the distortion caused by the difference in sensitivity between the fluorescent dyes, the distortion is described by means of a nonparametric regression model represented by the following EQ25, a measurement error caused by the difference in sensitivity between the fluorescent dyes is estimated by a nonparametric smoothing method represented by the following EQ26 and EQ27, and the error is corrected:

$$v_{ij}^{k} = \phi\left(u_{ij}^{k}\right) + \varepsilon_{ij}^{k} \quad , \varepsilon_{ij}^{k} = N\left(0, v^{2}\right) \quad (25)$$

$$\eta_{ij}^{k} = v_{ij}^{k} - \hat{\phi}\left(u_{ij}^{k}\right) \quad (26)$$

$$\hat{y}^k_{ij}(1) = \frac{1}{2}\left(u^k_{ij} + \eta^k_{ij}\right)$$
$$\hat{y}^k_{ij}(2) = \frac{1}{2}\left(u^k_{ij} - \eta^k_{ij}\right) \quad (27)$$

23. The cDNA microarray data correction method according to claim 13, wherein, supposing that a probability of gene expression is lower than 0.5, it is assumed for the correction that the fluorescence intensity detected at more than half of the spots within each grid indicates a background noise or a systematic error.

24. The cDNA microarray data correction method according to claim 23, wherein, supposing that $L_k(c)$ and $M_k(c)$ indicate 25 and 50 percent points of the fluorescence intensity obtained in at least two gene expression intensity data channels in a grid, it is further assumed for the correction that $L_k(c)$ and $Mk(c) - Lk(c)$ are equal among the grids and the channels on condition that most genes are in a non-expression state and that a distribution of 50 percent point or lower of the fluorescence intensity is common to all grids and channels.

25. The cDNA microarray data correction method according to claim 23, wherein, denoting that $A_k(c)$, $L_k(c)$ and $M_k(c)$ indicate 35, 10 and 90 percent points of the fluorescence in a grid k for channel c, it is assumed for the correction that $A_k(c)$ and $M_k(c)-L_k(c)$ are common to all grids and channels.

26. A cDNA microarray data correction program for use in correcting global and local distortions of microarray data more precisely and correcting measurement errors caused by a difference in sensitivity between fluorescent dyes with a computer to execute the steps of:

inputting previously-adjusted gene expression intensity data, considering flag information indicating a removal of background noise and reliability of each spot;

standardizing the gene expression intensity data by using grid-by-grid order statistics for the input gene expression intensity data on condition that most genes are in a non-expression state;

outputting the standardized gene expression intensity data;

estimating a distortion depending on the spot position on grid coordinates for the standardized gene expression intensity data by a nonparametric smoothing method and correcting the data distortion depending on the spot position;

outputting the first corrected gene expression intensity data whose distortion depending on the spot position has been corrected;

performing an S-D transformation for the first corrected gene expression intensity data, estimating a potential distortion caused by a difference in sensitivity between the fluorescent dyes in the gene expression intensity data by the nonparametric smoothing method, and correcting the distortion caused by the difference in sensitivity between the fluorescent dyes; and

outputting the second corrected gene expression intensity data whose distortion caused by the difference in sensitivity between the fluorescent dyes has been corrected.

27. A computer-readable memory medium containing a cDNA microarray data correction program for use in correcting global and local distortions of microarray data more precisely and correcting measurement errors caused by a difference in sensitivity between fluorescent dyes with a computer to execute the steps of:

inputting previously-adjusted gene expression intensity data, considering flag information indicating a removal of background noise and reliability of each spot;

standardizing the gene expression intensity data by using grid-by-grid order statistics for the input gene expression intensity data on condition that most genes are in a non-expression state;

outputting the standardized gene expression intensity data;

estimating a distortion depending on the spot position on grid coordinates for the standardized gene expression intensity data by a nonparametric smoothing method and correcting the data distortion depending on the spot position;

outputting the first corrected gene expression intensity data whose distortion depending on the spot position has been corrected;

performing an S-D transformation for the first corrected gene expression intensity data, estimating a

potential distortion caused by a difference in sensitivity between the fluorescent dyes in the gene expression intensity data by the nonparametric smoothing method, and correcting the distortion caused by the difference in sensitivity between the fluorescent dyes; and

outputting the second corrected gene expression intensity data whose distortion caused by the difference in sensitivity between the fluorescent dyes has been corrected.